Big Picture

Current distribution / Prior (using previous poll, and a choice of model: beta distribution)

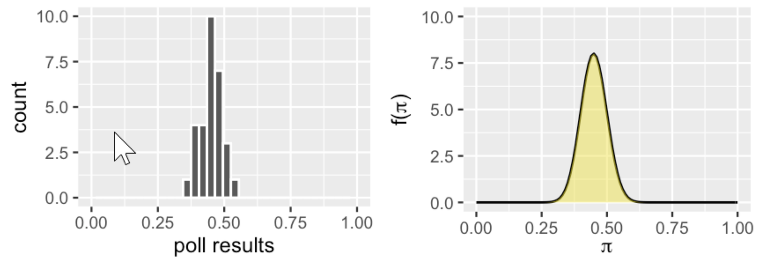
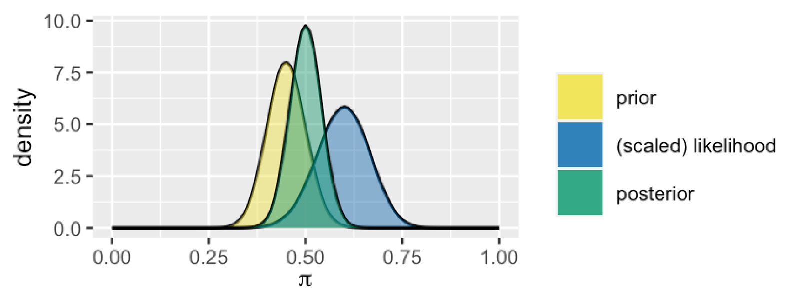


FIGURE 3.1: The results of 30 previous polls of Minnesotans' support of Michelle for president (left) and a corresponding continuous prior model for π, her current election support (right).

Data: New Poll

In *reality*, we ultimately observe that the poll was a huge success: $Y = 30$ of $n = 50$ (60%) polled voters support Michelle! This result is highlighted by the black lines among the pmfs in Figure 3.4. To focus on

Updated distribution using Bayesian principles



[Example of a Bayesian posteiror update analytically. Source: https://www.bayesrulesbook.com/chapter-2.html](https://www.bayesrulesbook.com/chapter-2.html)

Details

Find best fitting distribution parameters (fit visually or by some metric)

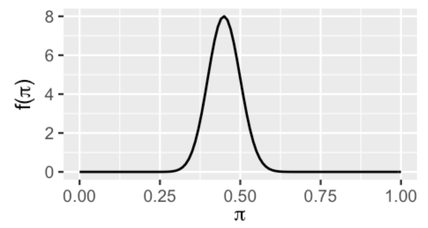


FIGURE 3.3: The Beta(45,55) probability density function.

Thus, a *reasonable* prior model for Michelle's election support is

$$\pi \sim \text{Beta}(45, 55)$$

The Beta model

Let π be a random variable which can take any value between 0 and 1, i.e., $\pi \in [0,1]$. Then the variability in π might be well modeled by a Beta model with **shape hyperparameters** $\alpha > 0$ and $\beta > 0$:

$$\pi \sim \text{Beta}(\alpha, \beta).$$

The Beta model is specified by continuous pdf

$$f(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1} \quad \text{for } \pi \in [0,1] \qquad (3.1)$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-y}dx$ and $\Gamma(z+1) = z\Gamma(z)$. Fun fact: when z is a positive integer, then $\Gamma(z)$ simplifies to $\Gamma(z) = (z-1)!$.

Bayes:
posterior ~ prior * likelihood

$$f(\pi|y=30) = \frac{f(\pi)L(\pi|y=30)}{f(y=30)}.$$

$$L(\pi|y=30) = \binom{50}{30}\pi^{30}(1-\pi)^{20} \quad \text{for } \pi \in [0,1].$$

hat we see in Figure 3.5, the chance that $Y = 30$ of 50 pol

if her underlying support were $\pi = 0.6$:

$$L(\pi = 0.6|y=30) = \binom{50}{30}0.6^{30}0.4^{20} \approx 0.115$$

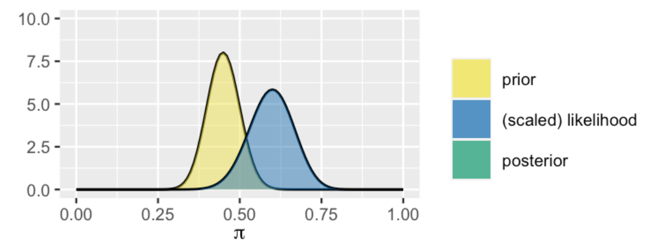
erlying support were $\pi = 0.5$:

$$L(\pi = 0.5|y=30) = \binom{50}{30}0.5^{30}0.5^{20} \approx 0.042.$$

Example likelyhoods for poll 30/50 using binomial

Calculated prior and likelihood distributions

$$Y|\pi \sim \text{Bin}(50, \pi)$$
$$\pi \sim \text{Beta}(45, 55).$$



$$\begin{aligned} f(\pi|y=30) &\propto f(\pi)L(\pi|y=30) \\ &= \frac{\Gamma(100)}{\Gamma(45)\Gamma(55)}\pi^{44}(1-\pi)^{54} \cdot \binom{50}{30}\pi^{30}(1-\pi)^{20} \\ &= \left[\frac{\Gamma(100)}{\Gamma(45)\Gamma(55)}\binom{50}{30}\right] \cdot \pi^{74}(1-\pi)^{74} \\ &\propto \pi^{74}(1-\pi)^{74} \end{aligned}$$

Multiply concretely parameterized binomial and beta

that is, the **kernel** of the pdf. Notice here that $f(\pi|y=30)$ has the *same* kernel as the normalized Beta(75,75) pdf in (3.9):

$$f(\pi|y=30) = \frac{\Gamma(150)}{\Gamma(75)\Gamma(75)}\pi^{74}(1-\pi)^{74} \propto \pi^{74}(1-\pi)^{74}.$$

Find best fitting distribution parameters (by solving for alpha/beta or finding the most fitting ones)

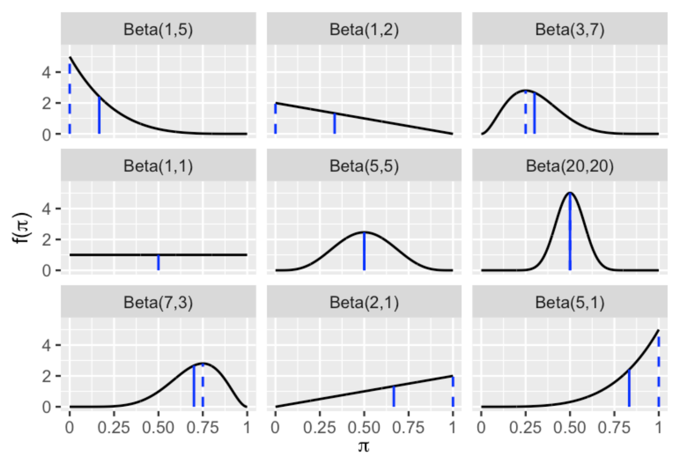


FIGURE 3.2: Beta(α, β) pdfs $f(\pi)$ under a variety of shape hyperparameters α and β (black curve). The mean and mode are represented by a blue solid line and dashed line, respectively.

The Binomial model

Let random variable Y be the *number of successes* in a *fixed number of trials* n. Assume that the trials are *independent* and that the *probability of success* in each trial is π. Then the conditional dependence of Y on π can be modeled by the Binomial model with **parameters** n and π. In mathematical notation:

$$Y|\pi \sim \text{Bin}(n, \pi)$$

where "$\sim$" can be read as "modeled by." Correspondingly, the Binomial model is specified by **conditional pmf**

$$f(y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y} \quad \text{for } y \in \{0,1,2,\ldots,n\} \qquad (2.7)$$

where $\binom{n}{y} = \frac{n!}{y!(n-y)!}$.

Beta Distribution

The two are conjugates (can be proven using this definition)

Binomial Distribution

Conjugate prior

Let the prior model for parameter θ have pdf $f(\theta)$ and the model of data Y conditioned on θ have likelihood function $L(\theta|y)$. If the resulting posterior model with pdf $f(\theta|y) \propto f(\theta)L(\theta|y)$ is of the same model family as the prior, then we say this is a **conjugate prior**.